APPENDIX C

DATA ADMINISTRATION TOOLS

A. GENERAL

1. INTRODUCTION

a. This Appendix, provides information for selecting and using tools to support DoD Data Administration. The DoD Data Administration infrastructure is not linked to, or supported by, a particular commercial or internally developed information or software engineering methodology; however, it will support the use of commercial methodologies. Selection and use of any particular methodology will not be static. In accordance with FIP PUB 184, "Integration Definition for Information Modeling (IDEF1 X)" (reference (r)), IDEF1X shall be used to integrate data models.

b. There are a variety of views of the scope of data administration and the methods to perform it. Some focus purely on the data in analysis and modeling; other methodologies recognize a strong link between activity and data analysis and modeling. Others say that data and activities should be integrated into an object-oriented approach. Whatever methodology is selected, it must emphasize the DoD-wide view of data, information system development, utilization and evolution of data and activities in existing systems, and collective integration of DoD data resources.

2. TOOL CAPABILITIES

a. To date, there are a number of computer-aided software engineering (CASE) tools that support the full life-cycle, from analysis to code generation. The later system development phases, such as, system implementation and physical database design are supported by only a few, highly specialized tools.

(1) The majority of CASE tools support the early systems development phases, such as Strategic Systems Planning and Requirements Analysis. Such tools are popularly called front-end tools.

(2) The mid-level system development phases, such as System Design and Logical Database Design, are supported by a broad number of specialized CASE tools.

(3) The later system development phases, such as System Implementation and Physical Database Design, however, are supported by only a few, highly specialized tools. These tools, including code generators and fourth generation languages, are called back-end tools.

b. At a minimum, appropriate data administration tools should provide support for the following functions:

(1) Project management.

(2) Data dictionary services.

(3) Activity and data modeling tools.

(4) CASE tools.

(5) Re-engineering and reverse engineering.

(6) Database management.

(7) Tools for data analysis, cross-referencing, and mapping of data in existing systems.

(8) Data quality assurance tools.

(9) Configuration management tools.

(1 O) Tools to assist in data synonym and redundancy control.

3. METHODOLOGIES TO SUPPORT DATA ADMINISTRATION IN THE DEPARTMENT OF DEFENSE

a. Linking methodologies and tools currently being used throughout the Department of Defense is essential to the successful development of the data administration infrastructure. This will aid in the development and modernization of the Department of Defense's automated and non-automated information systems. Components should attempt to migrate to Component-wide (and ultimately, DoD-wide) common methodologies and sets of tools (e. g., information engineering and CASE tools) that conform to the following criteria as closely as possible:

(1) Assist all DoD organizations to develop strategic, tactical, and operational data and activity models based on their mission, functions, goals, objectives, and enterprise strategies.

(2) Link strategic planning, modeling of the plan, and management of the implementation of the plan together as a coherent functional activity.

(3) Consist of a formal set of interrelated disciplines, procedures, techniques, deliverables, and quality assurance tests; and support the entire data and information system life-cycle for organization strategic functional planning, and requirements to data or information system replacement or retirement.

(4) incorporate integration of data and activity analysis and modeling that ideally will be capable of evolving to support emerging methodologies and techniques.

(5) Be capable of using the same techniques and procedures on any size or complexity of organization or project.

(6) Be easy to learn, understand, and use by executive, senior, and middle functional managers and personnel as well as data processing personnel because it will be used to identify the management and functional requirements of the organization.

(a) This includes the use of the methodology to assist management in the development and maintenance of strategic, as well as tactical and operational plans.

(b) This includes the production of data, activity, and possibly object models and reports that managers, users, and technicians can easily understand and use.

(7) Facilitate the documentation, analysis, refinement, and development of enterprise rules through models and reports.

(8) Be capable of defining the data architecture.

(9) Be capable of defining the functional partitions in the enterprise, analyzing those partitions to identify databases, developing database plans, and guiding and overseeing the function of database administration.

(1 O) Be rigorous enough to link each succeeding methodological phase to each other and ultimately to the technical or physical environment in which the logical data models will be implemented.

(11) Control and document data redundancy and facilitate systems integration top-down.

(12) Maintain technological or physical independence throughout the logical stages of the methodology.

(13) Be capable of effectively supporting transition from the logical to the physical stages of the methodology.

(14) Support the creation, development, maintenance, and management of flexible information systems that can be rapidly and readily modified.

b. Having the capacity to provide an organization-wide, top-down, shared data, business planning strategy also is essential. The Department of Defense should have those capabilities in the CASE tools it procures and uses.

B. SPECIFIC

DoD Data Administration is supported primarily by the following tools:

1. Information Resource Dictionary System (IRDS) tools.

The fundamental tool to support Data Administration is the data dictionary. The terms data directory, data encyclopedia, and data repository are widely used in the same context as data dictionary, but no standard definitions are universally accepted. For the purposes of this Manual, all three support the concept of "a specialized database containing information about data, such as meaning, relationships to other data, origin, usage, and format, including the information resources needed by an organization. "

a. An "encyclopedia" is more frequently used to reference the dictionary and directory features of CASE tools.

b. A "repository" denotes more robust functionality and includes full extensibility, versioning, security, and other specific services such as data model notation and metadata storage, diagram generation, and related services.

c. A "data directory" is defined by Federal Information Processing Standard (FIPS) 156 (reference (y)) as a subset of a data dictionary and/or directory that identifies data location and ownership.

(1) The DDRS is the data dictionary system which supports DoD Data Administration as specified in DoD Directive 8320.1 (reference (c)).

(.2) The current IRDS standard (reference (y)) was established to support the development of automated tools which in turn will support the application of data administration data standards and procedures. The standard assumes no implementation environment and assumes no run-time or compile-time dependent interfaces. The DDRS will become conformant with the FIPS 156 (reference (y)) as the IRDS standard at the earliest possible date.

(3) Component and Functional Area data dictionaries will be logically integrated and consistent with the DDRS but may be separately implemented apart from the DDRS. At a minimum, the DDRS will serve as the "dictionary and clearinghouse" for DoD standard data elements that all DoD Components and Functional Areas will be sharing to control and reduce redundancy and improve efficiency of database operations. The DDRS will evolve to support the more robust functionality.

d. The DDRS will:

(1) Support metadata definition, description, and management including the cross-referencing of information.

(2) Include schema extensibility, metadata analysis, and easy-to-use report generation.

(3) Provide guidance to users on how to follow one or more of the selected commercial or developed methodologies in support of DoD Data Administration.

(4) Provide support of users for metadata naming analysis and verification.

(5) Provide a standard method and functionality to support metadata interchange.

2. Computer-aided Software Engineering (CASE) tools.

a. CASE tools are most often used to assist in analyzing and designing information systems. CASE tools provide a data element consistency check against the entries in the DDRS. These tools generally include graphical support for the activity-orientated functional requirements decomposition methods, and the data-centered information engineering methods.

b. Support for configuration management is provided by maintaining the traceability links from the data models to the functional requirements being satisfied by the application software under development. The software life-cycle can be supported by maintaining the actual application software program listings and generated code in the metadata dictionary along with the required documentation. When a fully integrated CASE tool is utilized in this manner, the functional processing requirements, data, models, database schema, software, and documentation are available electronically for reuse or modification. This facilitates the life-cycle maintenance of the requirements.

3. Other Tools.

There are other tools (whether considered CASE or not) that might assist in supporting the development of a successful infrastructure:

a. A data element creation or assistance tool (generally custom built); i.e., if the data dictionary cannot check the data element naming conventions, a separate automated tool might be designed or acquired for that purpose.

b. Reverse engineering and re-engineering data modeling tools. Reverse

engineering tools can be used to develop data models where none exist.

c. Data-related quality assurance tools.

d. Fourth generation language and code generating tools.